Presentations

Alena Anishchanka, Natalia Levshina, Dirk Speelman, Dirk Geeraerts
University of Leuven, QLVL (Quantitative Lexicology and Variational Linguistics)
Alena.Anishchanka@arts.kuleuven.be, Natalia.Levshina@arts.kuleuven.be,
Dirk.Speelman@arts.kuleuven.be, dirk.geeraerts@arts.kuleuven.be

Shades of salience: Multivariate analysis of prototypicality effects in color terms

Research question

The proposed study aims to develop a rigorous quantitative method to measure the prototypicality effects in the semantics of lexical units, using color terminology in American advertising as the source of data. We approach this goal from two perspectives. Firstly, we analyze the psycholinguistic techniques developed in basic color categorization studies for measuring salience (basicness) of color terms and compare them to corpus-based measurements of color term salience. Secondly, we analyze a number of color terms using a bottom-up quantitative procedure for measuring their relative salience based on formal linguistic properties. Following the usage-based approach to semantics, we specifically avoid the *a priori* distinction between more prototypical (basic) and less prototypical (non-basic) color terms. Using the convergent evidence from psycholinguistic and corpus-based measurements, we suggest that a number of techniques developed in corpus-driven studies could be added to the arsenal of measurements applied in the basic color term (BCT) tradition. This would allow extending the analyses and generalizations to include a wider range of color terms and categories.

Background

Over the last half century, anthropological, psycholinguistic and linguistic studies of color categorization have developed a number of quantitative experimental procedures for measuring the psychological and linguistic salience (basicness) of color terms (see, for instance, Berlin & Kay 1969, Rosch Heider 1972, Rosch & Mervis 1975, MacLaury 1997, Corbett & Davies 1997). Many of these techniques such as reaction time and consistency in (color) naming experiments, frequency and ranking in elicitation tasks, frequency, derivational potential and length in text corpora have been successfully applied to other semantic domains in search of prototype representations. However, in the color domain itself the discussed measurements have been mostly limited in their application to basic color term analyses.

Alternatively, the recent corpus-based studies of color terms demonstrate that their basicness or salience is a matter of degree and cannot be reduced to the dichotomy of basic vs. non-basic color terms (Kerttula 2002) and that it might be affected by contextual factors (Steinvall 2002). Furthermore, the comparison of different behavioral and linguistic tests (Corbett & Davies 1997) demonstrates that not all of them are equally successful in distinguishing between basic and non-basic color terms. These findings give evidence to an internal hierarchy and graded membership in color categories in line with the multivariate model of semantics developed in Cognitive

Linguistics (Geeraerts et al. 1994, Gries 2003, Glynn 2007, Geeraerts 2010). They might also suggest the heterogeneous nature of salience that includes linguistic, categorical, and cultural levels (Geeraerts 2006).

Data and variables

The study is based on an extensive self-compiled corpus of color names and color samples used by US manufacturers and retailers in online marketing in four product categories (automobiles, clothing, make-up, and house paints).

For the purposes of the presented analyses, we randomly selected a sample of 16400 observations equally representing the four product categories. The data were annotated for a range of quantified formal and corpus-based linguistic parameters intended to operationalize the salience of color terms (see Table 1). For the analysis of correlation between corpus-based and experimental parameters of salience we used measurements applied in psycholinguistic color naming and elicitation experiments (Boynton & Olson 1990, Corbett & Davies 1997, Taft & Sivik 1997, Sturges & Whitfield 1997) (see Table 1).

TABLE 1. MEASUREMENTS OF SALIENCE.

Formal linguistic characteristics
length of the color term in alphanumeric characters
length of the color term in constituents
length of the color term in phonemes
length of the color term in syllables
Corpus-based characteristics
token frequency
type token frequency ratio
usage as an independent color name
attraction to the head or modifier slot in compound color terms as a simple ratio
attraction to the head or modifier slot in compound color terms as a log-likelihood ratio
the size of the color category according to the distribution of its exemplars in rgb color space
(un)restricted distribution across different product categories
derivational productivity based on type frequency of derived color terms
derivational productivity based on token frequency of derived color terms
Psycholinguistic parameters in color naming experiments
response time in milliseconds
consistency of naming
Psycholinguistic parameters in color name elicitation experiments
relative frequency
sequence rank
Evolutionary parameters
position in the evolutionary BCT hierarchy suggested by Berlin and Kay

The results of statistical analyses

In the first part of the study, we have tested the relationships between corpus-based and experimental parameters of salience using Kendall tau rank correlation coefficient and hierarchical cluster analysis based on the data available for 34 color categories.

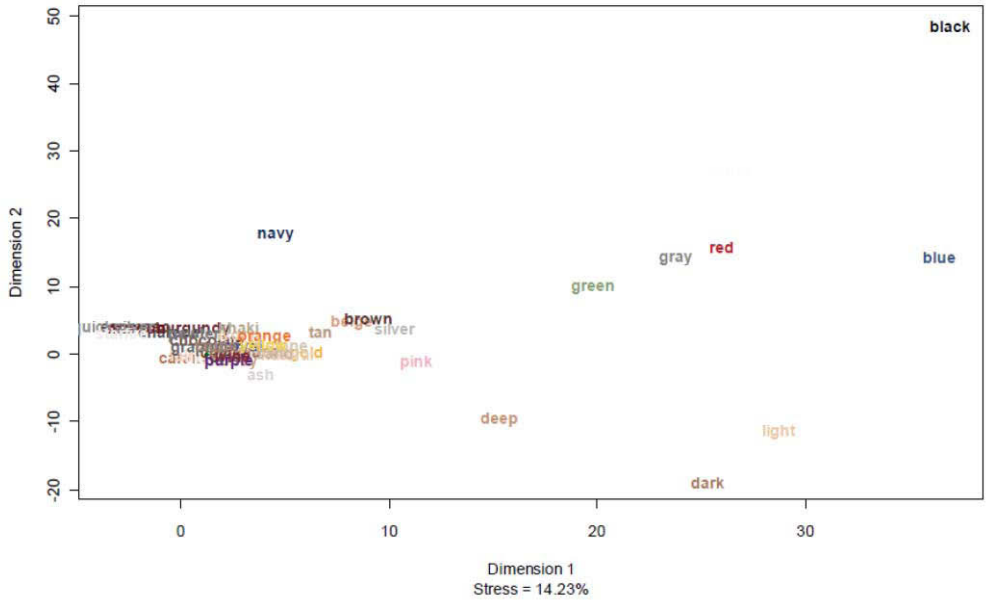
The results suggest several groups of salience parameters. The purely formal measurements of color term length show the highest correlation and form a distinct cluster. Corpus-based measurements are distributed in two clusters interrelated with other parameters. In one cluster several corpus-based characteristics including derivational productivity and usage in the head position correlate with Berlin and Kay

evolutionary sequence. In the other cluster such corpus-based measurements as token frequency and independent usage group with psycholinguistic measurements suggesting convergence of experimental and corpus evidence.

In the second part of the study, we performed a series of Kruskal's non-metric multidimensional scaling analyses exploring the relative salience of color terms used in advertising. The results confirm that basicness or salience of color terms is a continuous non-homogeneous parameter rather than a dichotomy. We can observe a continuum from the traditionally recognized BCT (*green, blue, red*) followed by the so-called secondary BCT (*orange, purple, pink*) through well entrenched non-basic monolexemic names (*wine, burgundy, charcoal)* to compounds with color modifiers (*light blue*) and finally to the most idiosyncratic compound terms (*deep sea blue, titanium silver*).

If we further zoom in on monolexemic color terms (see Figure 1), we can observe that most of the primary BCT (*black, blue, red, green*) are more distinct in their linguistic behavior compared to secondary BCT (*pink, orange, brown, purple*) and especially to very densely clustered non-basic color terms. At the same time, certain non-basic color terms (*silver, tan, beige*) come close in their linguistic characteristics to secondary BCT suggesting a continuum between basic and non-basic color terms.

FIGURE 1. CORPUS-BASED SALIENCE OF MONOLEXEMIC COLOR TERMS.



The proposed analyses make several contributions to the studies of linguistic salience and color categorization. Firstly, the comparison of corpus-based and experimental measurements of salience suggests both convergence of the two paradigms and the distinct place of corpus-based parameters. This can be seen as evidence of multidimensional nature of linguistic salience and prototypicality effects that create the basis for conceptual variation. Secondly, the chosen granularity of the analyses specifically addresses the gap between the categorization studies focusing on the most salient (basic) color categories and idiosyncratic color terms hand-picked for the studies of color terms in advertising (Stoeva-Holm 1996, Bergh 2007, Graumann 2007, Wyler 2007). Finally, considering the potentially unlimited source of data, the proposed model allows developing generalizations on a larger scale than has been suggested in the

previous basic color categorization analyses and the studies of color terms in advertising.

References

Bergh, Gunnar 2007. The semiosis of Swedish car color names: Descriptive and amplifying functions. In *Anthropology of Color: Interdisciplinary Multilevel Modeling,* MacLaury, Robert, Paramei, Galina and Dedrick, Don (eds.), 337-345. Amsterdam: John Benjamins.

Berlin, Brent and Kay, Paul 1969. *Basic Color Terms: Their Universality and Evolution.* Berkeley: University of California Press.

Boynton, Robert and Olson, Conrad 1990. Salience of chromatic basic color terms confirmed by three measures. *Vision Research* 30(9), 1311-1317.

Corbett, Greville and Davies, Ian 1997. Establishing basic color terms: measures and techniques. In *Color categories in thought and language,* Hardin, Clyde and Maffi, Luisa (eds.), 197-223. Cambridge: Cambridge University Press.

Geeraerts, Dirk 2006. Salience phenomena in the lexicon. A typology. In Geeraerts, Dirk *Words and Other Wonders. Papers on Lexical and Semantic Topics* [Cognitive Linguistics Research 33], 76-98. Berlin: Mouton de Gruyter.

Geeraerts, Dirk 2010. *Theories of lexical semantics*. Oxford: Oxford University Press.

Geeraerts, Dirk, Grondelaers, Stef and Bakema, Peter 1994. *The Structure of Lexical Variation. Meaning, Naming, and Context* [Cognitive Linguistics Research 5]. Berlin: Mouton de Gruyter.

Glynn, Dylan. 2007. Mapping meaning: Towards a usage-based methodology in Cognitive Linguistics. PhD dissertation, Katholieke Universiteit Leuven.

Gries, Stefan Th. 2003. *Multifactorial Analysis in Corpus Linguistics: The Case of Particle Placement* [Open Linguistic Series]: London/New York: Continuum Press.

Graumann, Andrea 2007. Color names and dynamic imagery. In *Speaking of colors and odors* [Converging Evidence in Language and Communication Research 8]*,* Plümacher, Martine and Holz, Peter (eds.), pp.129-140. Amsterdam: John Benjamins.

Kerttula, Seija 2002. *English Colour Terms: Etymology, Chronology, and Relative Basicness*. [Mémoires de la Société Néophilologique de Helsinki], Vol. LX. Helsinki: Société Néophilologique.

MacLaury, Robert 1997. Color and Cognition in Mesoamerica: Constructing Categories as Vantages. Austin: University of Texas Press.

Rosch Heider, Eleanor 1972. Universals in color naming and memory. *Journal of Experimental Psychology* 93(1), 10-20.

Rosch, Eleanor and Mervis, Carolyn 1975. Family resemblances: studies in the internal structure of categories. *Cognitive Psychology* 7, 573-605.

Steinvall, Anders 2002. English Color Terms in Context. PhD dissertation, Umeå University: Skrifter från moderna språk 3.

Stoeva-Holm, Dessislava 1996. Farbbezeichnungen in deutschen Modetexten. Eine morphologisch-semantische Untersuchung. [Studia Germanistica Upsaliensia, 34]. PhD dissertation, Uppsala: Universitetsförlag Almqvist & Wiksell.

Sturges, Julia and Whitfield, T.W.Allan 1997. Salient Features of Munsell Colour Space as a Function of Monolexemic Naming and Response Latencies. *Vision Research* 37(3), 307-313.

Taft, C. and Sivik, L. 1997. Salient color terms in four languages. *Scandinavian Journal of Psychology*, 38, 29-34.

Wyler, Siegfried 2007. Color terms between elegance and beauty: The verbalization of color with textiles and cosmetics. In *Speaking of colors and odors* [Converging Evidence in Language and Communication Research 8], Plümacher, Martine and Holz, Peter (eds.), pp.113-128. Amsterdam: John Benjamins.